

Red neuronal artificial para estimar un índice de calidad de agua

Artificial neural network to estimate an index of water quality

Lenin Quiñones Huatangari,¹ Luis Ochoa Toledo,² Nicolás Kemper Valverde,²
Oscar Gamarra Torres,³ José Bazán Correa,⁴ Jorge Delgado Soto¹

Resumen

La red neuronal artificial (RNA) es un modelo computacional que emula el sistema neuronal biológico en el procesamiento de información; los modelos que se originan son adecuados con el propósito de describir pronósticos para tiempos prolongados, además de relaciones no lineales. Se utiliza esta herramienta con el fin de predecir parámetros físico-químicos y microbiológicos que influyen en la calidad de agua. La Fundación Nacional de Saneamiento de Estados Unidos propuso un índice de calidad de agua, conocido como NSF WQI. Este artículo describe el diseño, entrenamiento y uso del modelo de la red neuronal perceptrón de tres capas para el cálculo del NSF WQI del río Utcubamba y sus afluentes. Empleando el *software* Matlab y aplicando el algoritmo de entrenamiento de Levenberg-Marquardt, se encontró que la arquitectura óptima de la RNA es 6-12-1, además, el porcentaje para los conjuntos de entrenamiento, validación y prueba fue de 70 %, 10 % y 20 %, respectivamente. El rendimiento de la RNA se ha evaluado utilizando la raíz del error cuadrático medio (RMSE) y el coeficiente de correlación (R). Se mostraron correlaciones altas (mayores que 0.94) entre los valores medidos y predichos. Finalmente, la RNA propuesta ofrece una alternativa útil para el cálculo y la predicción del índice de calidad de agua en relación con el oxígeno disuelto (OD), la demanda bioquímica de oxígeno (DBO), nitratos, coliformes fecales, potencial de iones hidrógeno (pH) y la turbidez.

Palabras clave

Índice de calidad de agua; redes neuronales artificiales; perceptrón multicapa; parámetros físico químicos.

Abstract

The artificial neural network (RNA) is a computational model that emulates the biological neural system in information processing. The originating models are suitable for the purpose of describing long-term specifics, in addition to nonlinear relationships. This tool is used to predict physical chemical and microbiological parameters that influence water quality. The United States National Sanitation Foundation proposed a water quality index, known as the NSF WQI. This article describes the design, training and use of the three-layer neural perceptron neural model for the calculation of the NSF WQI of the Utcubamba River and its tributaries. Using the Matlab software and applying the Levenberg-Marquardt training algorithm, the optimal RNA architecture was found to be 6-12-1, plus the percentage for the training, validation, and test sets of 70 %, 10 %, and 20 % respectively. RNA performance has been evaluated using the root of the root mean square error (RMSE) and the correlation coefficient (R). High correlations (greater than 0.94) were made between the measured and predicted values. Finally, the RNA proposal offers a useful alternative for the calculation and prediction of the water quality index in relation to dissolved oxygen (DO), biochemical demand for oxygen (BOD), nitrates, fecal coliforms, potential for hydrogen ions (pH) and turbidity.

Keywords

Water quality index; artificial neural networks; multilayer perceptron; physical-chemical parameters.

1 Universidad Nacional de Jaén, Jaén, Perú (lenin.quinones@unj.edu.pe, jorge.delgado@unj.edu.pe).

2 Universidad Nacional Autónoma de México, Ciudad de México, México (luis.ochoa@icat.unam.mx, nicolas.kemper@icat.unam.mx).

3 Universidad Nacional Toribio Rodríguez de Mendoza, Chachapoyas, Perú (oscar.gamarra@untrm.edu.pe).

4 Universidad Nacional de Piura, Piura, Perú (jbazanc@yahoo.com).

1. Introducción

Las aguas superficiales están sometidas a contaminación natural (arrastre de material particulado y disuelto, además de la presencia de materia orgánica natural) y de origen antrópico (descargas de aguas residuales domésticas, escorrentía agrícola, efluentes de procesos industriales, entre otros). El deterioro de las fuentes de abastecimiento de agua incide directamente en el nivel de riesgo sanitario presente y en el tipo de tratamiento requerido para su reducción, por lo que la evaluación de la calidad del agua permite tomar acciones de control y mitigación del mismo, garantizando el suministro de agua segura. Una herramienta común son los índices de calidad de agua (ICA), existe información de que en Alemania, en 1848, ya se realizaban algunos intentos por relacionar la presencia de organismos biológicos con la pureza del agua. En los últimos 130 años, varios países europeos han desarrollado y aplicado diferentes sistemas para clasificar la calidad de las aguas (Torres, Cruz y Patiño 2009).

La RNA es un modelo computacional capaz de imitar las características básicas del cerebro humano como la autoadaptabilidad, autoorganización y la tolerancia a errores. Durante las últimas dos décadas, sus aplicaciones han crecido de manera exponencial en casi todas las áreas de investigación, debido a que permiten resolver problemas que son susceptibles a la estadística convencional, por consiguiente, los modelos que se originan son adecuados para describir pronósticos para tiempos más prolongados, además de relaciones no lineales, por ende, es difícil prescribir la relación matemática exacta entre los parámetros. Se puede resumir, que han sido ampliamente adoptadas para la identificación, análisis, pronóstico, reconocimiento del sistema y optimización del diseño de modelos (López y Fernández 2008).

La RNA perceptrón multicapa (MLP) es un tipo de red, en el que su arquitectura consta de diversas capas, de nodos o neuronas interconectados, cada una de los cuales está conectado a todas las neuronas en la capa siguiente (ver figura 1). La capa de entrada, es en la cual se presentan los datos a la red neuronal, mientras que la capa de salida contiene la respuesta de la red neuronal. Además, posee capas intermedias, denominadas capas ocultas, estas pueden existir entre la capa de entrada y la capa de salida para permitir que estas redes representen y computar complicadas asociaciones entre patrones. Todas las neuronas ocultas y de salida procesan sus entradas multiplicando cada entrada por su peso, sumando el producto, y luego procesando la suma usando una función de transferencia no lineal para generar un resultado. Entre otros, el sigmoide es una de las funciones de transferencia más utilizada (López y Fernández 2008).

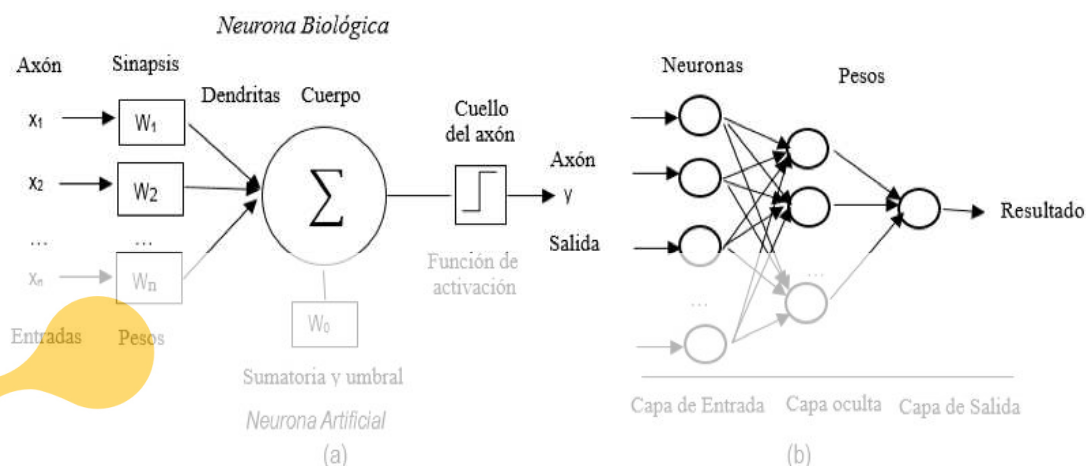
Entrenar una RNA es buscar un conjunto de pesos asociados a cada neurona con el fin de que la red pueda, a partir de datos de entrada, generar una salida; en el caso del aprendizaje supervisado, se tiene un conjunto de datos, pero no se conoce la función o relación matemática que los representa. Este proceso es iterativo, en el cual se va refinando la solución hasta alcanzar un nivel de operación suficientemente bueno (Gómez Rojas et al., 2004). La mayoría de los métodos de entrenamiento utilizados en las redes neuronales con conexión hacia delante consisten en proponer una función de error que mida el rendimiento actual de la red en función de los pesos. El objetivo del método de entrenamiento es encontrar el conjunto de pesos que minimizan (o maximizan) la función. El método de optimización proporciona una regla de actualización de los pesos que en función de los patrones de entrada modifica iterativamente los pesos hasta alcanzar el punto óptimo de la red neuronal (Federico Bertona, 2005). En la literatura especializada, encontramos diversos algoritmos con capacidad de entrenar redes neuronales:

SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark

descenso del gradiente, método de Newton, gradiente conjugado, cuasi-Newton, Levenberg-Marquardt, etc. (Sancho Caparrini, 2017).

Figura 1. (a) Similitudes entre la red neuronal biológica y RNA. (b) Red perceptrón multicapa



En este sentido, el empleo de las redes neuronales artificiales en el campo de recursos hídricos se ha vuelto cada vez más usual, es así que se ha propuesto un protocolo sistemático para el desarrollo y la documentación de las RNA, el cual se utilizó para revisar críticamente la calidad de los procesos de desarrollo e informes de modelos de RNA empleados en 81 artículos de revistas desde 2000, en los que la herramienta se ha utilizado para modelar la calidad del agua potable (Wu, Dandy y Maier 2014). Otros trabajos relevantes que se expresan brevemente son: se ha empleado RNA para predecir y pronosticar parámetros físico-químicos y microbiológicos (Wu, Dandy y Maier 2014).

Register for free at <https://www.scipedia.com> to download the version without the watermark

Para predecir la calidad del agua en Hilo Bay, Océano Pacífico, donde se aplicaron diferentes combinaciones de parámetros de calidad del agua como parámetros de entrada para predecir los valores diarios de salinidad, temperatura y oxígeno disuelto (Alizadeh y Kavianpour 2015). Modelos de predicción fueron desarrollados para alcalinidad, pH, calcio, dióxido de carbono, temperatura, dureza total, turbidez, sólidos disueltos totales y conductividad eléctrica tienen una capacidad de generalización aceptable, estos modelos pueden ser utilizados por operadores de planta de tratamiento de agua y tomadores de decisiones (Solaimany-Aminabad, Maleki y Hadi 2013). Sin embargo, Thambavani D y Uma (2014) describen el diseño y la aplicación del modelo de MLP de tres capas con alimentación directa y totalmente conectado para calcular el índice de calidad del agua (WQI) para Batlagundu, Dindigul District, Tamilnadu en la India, la arquitectura de red óptima fue 8-3-1, los parámetros de entrada fueron el pH, OD, EC, TDS, TA, TH, Ca, Mg y el algoritmo de entrenamiento fue el de descenso de gradiente. Las RNA se vienen empleando para estimar algún parámetro de calidad de agua de ríos, en función de otros parámetros físico-químicos, ver tabla 1.

El objetivo del presente trabajo de investigación fue emplear las redes neuronales artificiales para estimar el índice de la calidad del agua. La red neuronal propuesta emplea herramientas computarizadas, además nos faculta a usar razonamientos aproximados con base en información de campo y previa coordinación con expertos.

Tabla 1. Diseño, entrenamiento e implementación de redes neuronales artificiales perceptrón multicapa empleadas para estimar parámetros de calidad de agua de ríos

Diseño de la red			Parámetros		Implementación	Referencia
Algoritmo entrenamiento	Función de activac	#capas/Neuronas	Entrada	Salida	País/río	
Levenberg-Marquardt	Lineal a tramos	3 capas (11-23-1)	pH, ST, T-Alk, T-Hard, Cl, PO ₄ , K, Na, NH ₄ N, NO ₃ N, COD	OD	India/Gomti	(Singh et al., 2009)
	Tangsig	3 capas (11-11-1)		DBO		
Regla generalizada de aprendizaje delta	Tangente hiperbólica o sigmoidea ($\tanh(x)$)	4 capas (141-90-30-1)	Salinidad, flujo y nivel en diferentes puntos del río	Salinidad	Australia/Murray	(Maier y Dandy, 1996)
Descenso de gradiente	Lineal a tramos	4 capas (1-25-4-1)	Conductividad	SDT	Malasia/Johor	(Najah et al., 2009)
		4 capas (1-25-4-1)	Sólidos Disueltos Totales	Cond		
		3 capas (1-3-1)	Sólidos Totales Suspendidos	Turb		
Filtros de Kalman	Tangente hiperbólica o sigmoidea ($\tanh(x)$)	3 capas (9-26-1)	ECw, OD, Q, T, SO ₄ ²⁻ , Na ⁺ , +K ⁺ , TP, pH, NH ₄ ⁺	Nitratos	Grecia/Strymon	(Diamantopoulou et al., 2005)
		3 capas (7-8-1)	Q, Sulfatos, Na ⁺ , +K ⁺ , Mg ²⁺ , Ca ²⁺ , Cl ⁻ , HCO ₃ ⁻	Conduct Eléctrica	Grecia/Strymon	
		3 capas (10-10-1)	NO ₃ ⁻ , ECw, Q, T, SO ₄ ²⁻ , Na ⁺ , +K ⁺ , TP, pH, Cl ⁻ , NH ₄ ⁺	OD	Grecia/Strymon	
Descenso de gradiente	Logaritmo sigmoideo (logsig(x))	3 capas (8-3-1)	COD, HN ₃ -N, CL-A, NO ₂ -N, T, NO ₃ -N, OD, Qw	DBO	Turquía/Melen	(Dogan et al., 2009)
Levenberg-Marquardt	Sigmoidea Tangente hiperbólica	3 capas (4-10-2)	TSS, TS, pH, T	DBO, COD	Iran/Tehran	(Zare Abyaneh, 2014)

Las razones que han motivado a realizar el presente estudio son tres: la primera, desde el punto de vista científico es relevante la validación de un modelo basado en una RNA para estimar un índice de calidad de agua frente a la realidad física que se ha obtenido en estudio

previo en la cuenca del río Utcubamba. Segunda, desde el punto de vista funcional, al tener una herramienta computacional confiable en tiempo real para estimar la calidad del agua del río Utcubamba, las instituciones del Estado puede emplearla para realizar operaciones o proponer políticas ambientales que permitan mejorar la calidad del agua, y finalmente, la tercera es emplear modelos de minería de datos, por razones de facilidad de implementación, con un costo relativamente inferior al de los matemáticos como el del índice de calidad de agua NSF WQI, el mismo que necesita más parámetros.

2. Materiales y métodos

Metodología

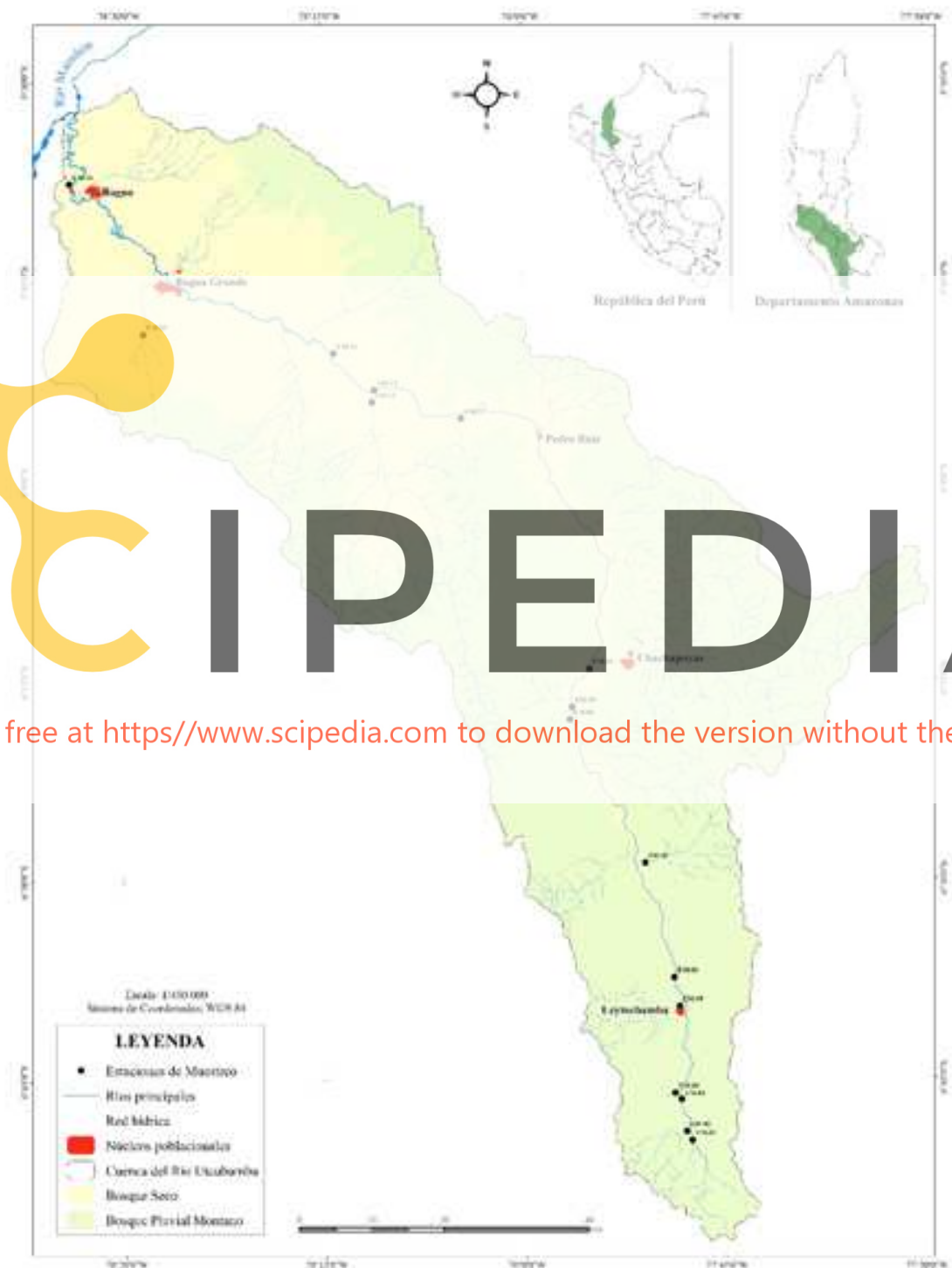
La metodología empleada fue de tipo analítica y consistió en tres fases: (1) Se recolectó y excluyó data sobre parámetros concerniente a calidad de agua del Río Utcubamba, del Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES-CES) de la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas. (2) Se implementó un modelo que estimó un índice de calidad de agua para el río Utcubamba de la región Amazonas basado en RNA; para este fin se empleó el *software* Matlab. (3) Finalmente, se validó el modelo, la cual consistió en comparar los residuos obtenidos entre los datos dados por el modelo y los datos recolectados en campo.

Base de Datos para estimar un índice de calidad de agua

Se realizó una base de datos tomando en cuenta los resultados de los parámetros de calidad ecológica del agua en la cuenca del río Utcubamba, Amazonas, Perú; reportados en la publicación del INDES-CES. Para este estudio previo, la recolección de las muestras de agua se realizó con la estación de lluvias (a lo largo del mes de octubre de 2014), se establecieron cuarenta y tres (43) estaciones de muestreo a lo largo de la cuenca principal y en los tributarios más importantes e influentes en la misma. Además, para la recogida de muestras para el análisis físico químico (15 parámetros) y microbiológico (4 parámetros) se siguió el protocolo de monitoreo de la calidad sanitaria de los recursos hídricos superficiales, establecido por la Dirección General de Salud Ambiental del Perú, mientras que para el análisis de las parámetros físico químicos y microbiológicas se hizo lo propio con los procedimientos dados en el *Standard Methods for the Examination of Water and Wastewater* (Gamarra et al., 2018).

La base de datos creada tuvo dos fines, el primero fue determinar el índice de calidad de agua y el segundo fue entrenar, validar y emplear la red neuronal artificial propuesta. Estuvo constituida por 16 instancias concernientes a la misma cantidad de puntos de muestreo; los mismos que fueron considerados debido al tipo de dato ordinal de los diez parámetros tomados en cuenta en el presente estudio, debido a que, para estimar el índice de calidad de agua, se necesitaba este tipo de valores. Cada instancia constó de diez (10) valores, siendo estos: temperatura, oxígeno disuelto, DBO₅, sólidos disueltos totales, fosfatos, nitratos, pH, coliformes fecales y turbiedad. En la figura 2, se puede observar la zona de estudio y los dieciséis puntos de muestreo.

Figura 2. Zona de estudio y 16 puntos de muestreo considerados sobre la cuenca del río Utcubamba.



SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark

El índice de la Fundación Nacional de Saneamiento (NSF WQI)

En Estados Unidos, la Fundación Nacional de Saneamiento (NSF, por sus siglas en inglés) desarrolló el NSF WQI en 1970 mediante el uso de la técnica de encuesta Delphi. Este índice tiene la característica de ser un índice multiparamétrico basado en tres estudios, donde cada una de los nueve parámetros tiene un peso específico de acuerdo con su importancia, relacionada a la calidad del agua. Estos pesos se muestran en la tabla 2, los cuales son aplicados con los parámetros para generar la media ponderada que constituye el ICA (Behar et al., 1997).

El índice NSF WQI fue calculado mediante la ecuación (1), con el propósito de probar la precisión de la red neuronal artificial para estimar el índice de calidad, por lo tanto, en términos de minería de datos, estos datos obtenidos vendrían a ser la verdad absoluta (*ground truth*).

Tabla 2. Pesos relativos a los parámetros que conforman el NSF WQI

Parámetro	NSF WQI
Temperatura	0.10
Oxígeno disuelto	0.17
DBO5	0.10
Sólidos Disueltos Totales	0.08
Turbiedad	0.08
Fosfatos	0.10
Nitratos	0.10
pH	0.12
Coliformes fecales	0.15

Si denotamos por I_i al subíndice correspondiente a la variable ' I ' y por w_i su ponderación respectiva, el ICA es expresado por:

$$ICA = \sum_{i=1}^9 w_i I_i \quad (1)$$

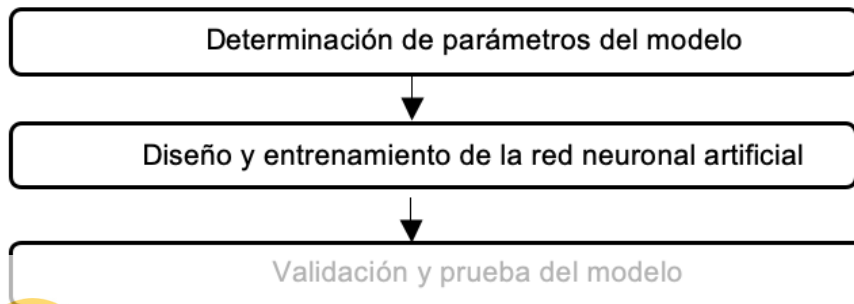
La clasificación de la calidad del agua, empleando el ICA, las cuales se expresan en cinco categorías, dependen del rango numérico dado: Muy mala (0-25), mala (26-50), media (51-70), buena (71-90) y excelente (91-100). Para el cálculo de este índice se utilizó la herramienta de forma online (Water Research Center, 2018).

Procedimiento para determinar la arquitectura y validación de la RNA

Al emplear una red neuronal, se consideran tres etapas. La primera consiste en el diseño, en el que se elige el tipo de red neuronal a utilizar (la arquitectura o topología), el número de neuronas que la compondrán, función de activación y el algoritmo de entrenamiento. La etapa de entrenamiento, en donde se le presenta a la red neuronal una serie de datos de entrada y datos de salida, para que a partir de ellos pueda aprender, gracias al uso del algoritmo de entrenamiento y, finalmente, el uso que consiste en suministrarles las entradas pertinentes de la red donde esta genera las salidas en función de lo que ha aprendido en la fase de entrenamiento (Dogan

et al., 2009). El procedimiento empleado para determinar la arquitectura y validación de la red neuronal artificial, consideró estas etapas, ver figura 3.

Figura 3. Procedimiento para determinar la arquitectura y validación de la RNA



Determinación de parámetros del modelo: siete parámetros se han contemplado en la presente investigación. Seis parámetros físico-químicos, los cuales son: oxígeno disuelto y la DBO, debido a que son parámetros químicos de la calidad del agua, nitratos, que es el parámetro que proporciona la información de nutrientes del agua. Además de los coliformes fecales, que es el parámetro biológico que determina el grado de contaminación por microorganismos termo tolerantes, por ende, parte de la calidad sanitaria del agua y, por último, el pH y la turbidez, ya que son parámetros físicos de la calidad del agua, estos parámetros (Srivastava et al., 2013) los propone para estimar un índice de calidad de agua mediante lógica difusa. Por otro lado, otro parámetro contemplado fue el NSF WQI. A fin de evitar impactos negativos en el empleo de la red neuronal es necesario durante la preparación inicial de los datos, estandarizar los datos, es decir, convertir los datos a una forma no dimensional de rango de variabilidad uniforme (Dawson y Wilby, 2001).

La cantidad de instancias de la base de datos disponible para modelar fue de dieciséis, que se asignaron de forma aleatoria en tres conjuntos, denominados de entrenamiento, validación y prueba, la distribución de estos es el 70 %, 10 % y 20 % de los datos ha sido propuesta en el trabajo de May y Sivakumar (2009).

Diseño y entrenamiento de la RNA. El fin fue predecir el estado de la calidad del agua del río Utcubamba (NSF WQI) en términos de seis parámetros físico-químicos, además de considerar los tipos de redes empleadas en estudios previos, ver tabla 1. Este estudio empleó una RNA supervisada y del tipo unidireccional denominado perceptrón de tres capas: La capa de entrada consta de seis parámetros físico-químicos, una capa oculta y el NSF WQI en la capa de salida.

La RNA almacena el conocimiento acerca del problema en términos de pesos en las interconexiones. El proceso de determinar los pesos a la RNA es denominado entrenamiento. La RNA es entrenada con un conjunto de entradas y salidas conocidas. Al principio del entrenamiento, los valores iniciales de los pesos son asignados aleatoriamente. Los pesos son sistemáticamente cambiados por el algoritmo de entrenamiento de forma tal que para una entrada la diferencia entre los resultados de la RNA y los datos actuales son pequeños. Un amplio número de criterios estadísticos están disponibles para comparar la bondad/suficiencia de algún modelo dado. La evaluación estadística del comportamiento usada para el entrenamiento en el presente trabajo fueron la raíz cuadrada media del error (*RMSE*) y el coeficiente de correlación (*R*). Estos valores han sido determinados empleando la ecuación (2) y ecuación (3) (Singh et al., 2009).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - q_i)^2}{n}} \quad (2)$$

$$R = \frac{\sum_{i=1}^n (Q_i - \bar{Q})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^n (Q_i - \bar{Q})^2 (q_i - \bar{q})^2}} \quad (3)$$

Donde $\bar{Q} = \frac{1}{n} \sum_{i=1}^n Q_i$, $\bar{q} = \frac{1}{n} \sum_{i=1}^n q_i$, Q son los valores observados y q los valores calculados.

Gazzaz et al. (2012) sostienen que el número de neuronas de la capa oculta debería estar dentro del intervalo $[\frac{l}{2}, 4l]$, donde l/O es el número de neuronas de la capa de entrada y salida, respectivamente. Por otro lado, Palani et al. (2008) han propuesto que estas neuronas no deberían ser de ninguna manera menor que el máximo de $\frac{l}{3}$ y 0. En nuestro caso, $l=6$ y $O=1$. Por ello, reemplazando los valores en las ecuaciones, se obtiene que el total de las neuronas ocultas se encuentra dentro del intervalo $[2, 24]$.

El procedimiento que se empleó para obtener la arquitectura óptima de la RNA, fue: (1) Entrenar tres corridas cada red de tres capas, donde las neuronas de entrada fueron 6, de salida 1 y las ocultas cada valor del intervalo $[2, 24]$, usando el Toolbox del *software* Matlab, cabe señalar que se utilizó únicamente el conjunto de entrenamiento. (2) Determinar el promedio de los resultados en las tres corridas de los estadísticos $RMSE$ y R , esto debido a que, en la etapa de entrenamiento, se utilizan datos en forma aleatoria para cada corrida. (3) Escoger el número de neuronas que compondría la capa oculta, bajo el criterio que el promedio de R (más cercano a uno) debe ser el mejor y el $RMSE$ el menor, con respecto a todos los promedios encontrados. Estas dos características deben cumplirse al mismo tiempo. (4) La arquitectura óptima de la RNA de tres capas, estuvo compuesta por seis parámetros en la capa de entrada, un parámetro en la capa de salida y la cantidad de neuronas determinadas en el paso (3) en la capa oculta.

En esta etapa, al obtener la arquitectura óptima la RNA es considerada entrenada.

Validación y prueba de la RNA: La validación de la RNA fue realizada mediante dos pasos: (1) Se ejecutaron tres corridas de la red con arquitectura óptima, usando el Toolbox del *software* Matlab, vale indicar que se utilizó únicamente el conjunto de validación. (2) Determinó el promedio de los resultados en las tres corridas de los estadísticos $RMSE$ y R , esto debido a que, al ejecutar las tres corridas, se utiliza datos en forma aleatoria para cada corrida.

El desempeño o prueba de la RNA se realizó de manera análoga al proceso de validación, con la particularidad que los datos donde se empleará el modelo serán en el conjunto prueba.

3. Resultados

La RNA construida fue entrenada usando el algoritmo de Levenberg-Marquardt, empleó la función de transferencia no lineal (tansig) se utilizó, tanto en la capa oculta como en la externa. Se determinó la distribución óptima, ver tabla 3, consistente de seis neuronas en la capa de entrada, doce neuronas en la capa oculta y una neurona en la capa de salida, el mismo que proporcionó un mejor modelo de ajuste para los tres conjuntos de datos. El coeficiente de correlación (R) para el entrenamiento, validación y los conjuntos de prueba fue de 0.979, 1 y 0.940,

respectivamente. Los valores respectivos de *RMSE* para los tres conjuntos de datos fue 2.562 para entrenamiento, 1.546 la validación y 1.997 la prueba.

Tabla 3. Desempeño de los parámetros de la RNA para estimar la calidad del agua del río Utcubamba (Perú)

Estructura RNA	Entrenamiento (70 %)		Validación (10 %)		Predicción (20 %)	
	RMSE	R	RMSE	R	RMSE	R
6-12-1	2.652	0.979	1.546	1	1.997	0.940

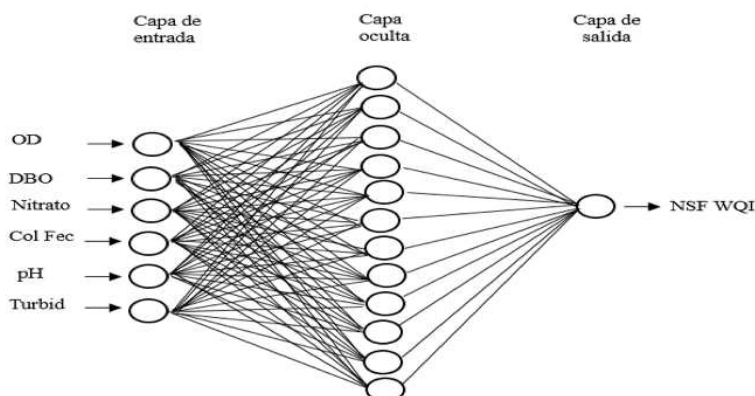
4. Discusión

Los datos que se han utilizado para diseñar, emplear y usar la red neuronal artificial, constó de dieciséis puntos de muestreo del río Utcubamba y sus afluentes, los mismos que fueron sustraídos del trabajo previo, realizado por Gamarra et al. (2018). La recolección de estas muestras de agua se realizó con la estación de lluvias (a lo largo del mes de octubre de 2014).

Se empleó la RNA del tipo perceptrón multicapa para estimar el índice de calidad de agua (NSF WQI), tomando como referencia los trabajos previos (ver tabla 1), considerando como parámetros de entrada a los parámetros: OD, DBO, nitratos, coliformes fecales, pH y turbidez. Diferentes modelos de redes neuronales fueron construidos y examinados para determinar el número óptimo de neuronas en la capa oculta, además de encontrar el porcentaje de distribución óptimo que correspondería a los conjuntos de entrenamiento, validación y pruebas para las muestras.

La arquitectura de la RNA óptima para estimar el índice de la calidad del agua de río (NSF WQI) se muestra en la figura 4. Con el trabajo de Thambavani D y Uma (2014), se coincidió en la aplicación del modelo de red neuronal perceptrón de tres capas para calcular el índice de calidad del agua. Sin embargo, como se mencionó en el trabajo de Thambavani D y Uma (2014), existen diferencias con el trabajo mencionado, como la arquitectura de red óptima (8-3-1), los parámetros de entrada, las cuales fueron el pH, OD, EC, TDS, TA, TH, Ca, Mg, el algoritmo de entrenamiento (retro propagación) y el WQI empleado es calculado bajo la normativa del país donde se realizó el estudio.

Figura 4. Red Neuronal Artificial para estimar el índice de calidad del agua del río Utcubamba (Perú)



Se determinó que la distribución óptima, obteniendo el valor de R mayores que 0.94 y el $RMSE$ menor que 2.65, para la distribución de los conjuntos de entrenamiento, validación y prueba de la RNA, es el 70 %, 10 % y 20 %; esta distribución es la misma encontrada en el trabajo de May y Sivakumar (2009).

5. Conclusiones y recomendaciones

En este trabajo, se propone un modelo basado en redes neuronales artificiales para el cálculo del índice NSF WQI, el mismo que permite determinar el índice de la calidad del agua del río Utcubamba y sus afluentes (Perú). La red neuronal artificial que se empleó ha sido el perceptrón de tres capas, el algoritmo de entrenamiento fue el de Levenberg Marquardt. El presente estudio muestra que la red neuronal artificial óptima que tiene la arquitectura 6-12-1 es capaz de capturar a largo plazo las tendencias observadas a la tediosa variable como es determinar el índice de la calidad del agua (NSF WQI), tanto en el tiempo como en el espacio.

Se destacan a las redes neuronales como una herramienta efectiva para el cálculo del índice de la calidad del agua del río Utcubamba, se concluye que estas podrían también ser utilizadas en otras áreas para mejorar la comprensión del río, tales como las tendencias de la contaminación. La RNA puede ser vista como un poderoso predictor, además de ser otra alternativa a las técnicas de modelado tradicionales.

Referencias

- Alizadeh, M. J., y Reza Kavianpour, M. (2015). Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine pollution bulletin*, 98 (1-2): 171-178.
- Behar, R., Zúñiga de Cardozo, Ma. C. y Rojas, O. (1997). Análisis y valoración del índice de calidad de agua (ICA) de la NSF: casos ríos Cali y Meléndez. *Revista Ingeniería y Competitividad*, 1 (1): 17-27.
- Chau, K. (2006). A review on integration of artificial intelligence into water quality modelling. *Marine Pollution Bulletin*, 52 (7): 726-733.
- Dawson, C. W., y R. L. Wilby. (2001). Hydrological Modelling Using Artificial Neural Networks. *Progress in Physical Geography: Earth and Environment*, 25 (1): 80-108.
- Diamantopoulou, M. J., Papamichail, D. M. y Antonopoulos, V. Z. (2005). The Use of a Neural Network Technique for the Prediction of Water Quality Parameters. *Operational Research*, 5 (1): 115-125.
- Dogan, E., Sengorur, B. y Koklu, R. (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management*, 90 (2): 1229-1235.
- Federico Bertona, L. (2005). Entrenamiento de redes neuronales basado en algoritmos evolutivos [Tesis de grado en Ingeniería Informática, Universidad de Buenos Aires]. <http://laboratorios.fi.uba.ar/lsi/bertona-tesisingenieriainformatica.pdf>
- Gamarra, O., Corroto, F., Rascón, J. y Chávez, J. (2018). *Calidad ecológica del agua en la cuenca del río Utcubamba, Amazonas, Perú (Primera)*. Primera. Chachapoyas: Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas.
- Gazzaz, N. M., Yusoff, M. K., Zaharin Aris, A., Juahir, H. y Firuz Ramli. M. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, 64 (11): 2409-2420.
- Gómez Rojas, G. A., Henao López, J. C. y Salazar Isaza, H. (2004). Entrenamiento de una red neuronal artificial usando el algoritmo simulated annealing. *Scientia Et Technica*, X: 13-18.

- Flórez López, R., Lévy Mangin, J. P. y Fernández Fernández, J. M. (2008). *Las redes neuronales artificiales*. Netbiblo: Tirant lo Blanch.
- Maier, H. R., y Graeme C. Dandy. (1996). The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research* 32 (4): 1013-1022.
- May, D., y Muttucumaru S. (2009). Prediction of urban stormwater quality using artificial neural networks. Faculty of Engineering - Papers (Archive): 296-302.
- Najah, A., Elshafie, A., Karim, O. A. y Jaffar, O. (2009). Prediction of Johor River water quality parameters using artificial neural networks. *European Journal of Scientific Research*, 28 (3): 422-435.
- Palani, S., Liong, S.-Y., y Tkalich, P. (2008). An ANN Application for Water Quality Forecasting. *Marine Pollution Bulletin*, 56 (9): 1586-1597.
- Sancho Caparrini, F. (2017). Entrenamiento de Redes Neuronales: Mejorando el Gradiente Descendiente. <http://www.cs.us.es/~fsancho/?e=165>.
- Singh, K. P., Basant, A., Malik, A. y Jain, G. (2009). Artificial neural network modeling of the river water quality—a case study. *Ecological Modelling*, 220 (6): 888-895.
- Solaimany-Aminabad, M, Maleki, A. y Hadi, M. (2013). Application of artificial neural network (ANN) for the prediction of water treatment plant influent characteristics. *Journal of advances in environmental health research*, 1 (2): 89-100.
- Srivastava, P., Burande, A. y Sharma, N. (2013). Fuzzy Environmental Model for Evaluating Water Quality of Sangam Zone during Maha Kumbh 2013. *Applied Computational Intelligence and Soft Computing*. Recuperado 24 de septiembre de 2018 <https://www.hindawi.com/journals/acisc/2013/265924/>.
- Thambavani D., S., y Uma, M. (2014). Numerical Study of Back Propagation Learning Algorithms for Forecasting Water Quality Index. *IJERST*, 3 (3): 1548-1555.
- Torres, P., Hernán Cruz, C. y Patiño, P. J. (2009). Índices de calidad de agua en fuentes superficiales utilizadas en la producción de agua para consumo humano. Una revisión crítica. *Revista Ingenierías Universidad de Medellín*, 8 (15 Sup. 1): 79-94.
- Water Research Center. (2018). Water Quality Index Calculator. Recuperado 9 de noviembre de 2018 (<https://www.water-research.net/index.php/water-treatment/water-monitoring/monitoring-the-quality-of-surfacewaters>).
- Wu, Wenyan, G., Dandy, C. y Maier, H. R. (2014). Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environmental Modelling & Software*, 54: 108-127.
- Zare Abyaneh, H. (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal of Environmental Health Science and Engineering* 12 (1): 40.